

A

221TCS100022303

Pages: 4

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

First Semester M.Tech Degree Regular and Supplementary Examination December 2023 (2022 scheme)



Discipline: COMPUTER SCIENCE AND ENGINEERING

Course Code & Name: 221TCS100 ADVANCED MACHINE LEARNING

Max. Marks: 60

Duration: 2.5 Hours

PART A

Answer all questions. Each question carries 5 marks

Marks

- 1 Let x_1, x_2, \dots, x_n be a random sample from a population with probability density function $f(x \vee \theta) = \frac{1}{\theta} x^{(1-\theta)/\theta}$, $0 < x < 1$, $0 < \theta < \infty$. Find the maximum likelihood estimate (MLE) for θ . (5)
- 2 The vectorized feature representation of an object, $\mathbf{x} = [2, 4, 1, 3]$ is given to a 3-layer neural network with one output layer and two hidden layers each having 3 neurons. Assume all inter-connection weights and bias having value 1. Use activation function $f(x) = 2x$ at hidden layers and **ReLU** at output layer. Draw schematic diagram of the network and compute the output for one round of forward propagation. (ReLU: $f(x) = \max(0, x)$) (5)

	A	B	C	D	E
A	0	1	2	2	3
B	1	0	2	4	3
C	2	2	0	1	5
D	2	4	1	0	3
E	3	3	5	3	0

- 3 Given the following distance matrix, construct the dendrogram using agglomerative clustering using complete linkage. (Hint: Complete linkage chooses the maximum distance when clusters are combined.) (5)
- 4 What is meant by kernel trick in the context of support vector machines? Show that the kernel function $K(\vec{x}, \vec{y}) = (\vec{x} \cdot \vec{y})^2$ is equivalent to $\phi(\vec{x})^T \phi(\vec{y})$, where $\phi(\vec{z}) = z_1^2 + z_2^2 + \sqrt{2}z_1z_2$, where each vector is represented in 2 dimensions. (5)

- 5 a) Consider a classifier trained using images of dogs and cats. Test set (X_{test}, Y_{test}) is such that the first m_1 images are of dogs, and the remaining images are of cats. After shuffling the test set, the model is evaluated and obtain a classification accuracy of $a_1\%$. Without shuffling the test set the model is again evaluated and obtain a classification accuracy of $a_2\%$. What is the relationship between a_1 and a_2 ? Explain. (2)
- b) Suppose two classifiers are trained to predict the probability of having cancer. First classifier attains 100% accuracy on the training set and 75% accuracy on test set. Second classifier attains 75% accuracy on the training set and 80% accuracy on test set. Which classifier you prefer? Justify (3)

PART B

Answer any 5 questions. Each question carries 7 marks

- 6 Predict the salary of a person having 5.5 years of experience using regression model generated from the following data, where salary is a function of experience. (7)

Experience (Years)	Salary (Thousand)
1	45
2	50
3	60
4	75
5	95
6	120
7	150

- 7 Explain the principle of the gradient descent algorithm. How does choosing a learning rate close to zero and choosing a learning rate close to one affect the performance of the Gradient Descent algorithm? (7)
- 8 The following table consists of training data from an employee database. Let "status" be the class label attribute. For a given row entry, "count" represents the (7)

number of data tuples having the values for department, status, age range, and salary range given in that row. Given a data tuple having the values "sales", "31...35" and "46...50k" for the attributes department, age range, and salary range respectively. Predict the status for the tuple using Naïve Bayes classifier.

Department	Status	Age Range	Salary Range	Count
Purchase	Senior	46....50	36k....40k	4
Purchase	Junior	26....30	26k....30k	10
Systems	Junior	21....25	46k....50k	20
Systems	Senior	31....35	66k....70k	5
Systems	Junior	26....30	46k....50k	3
Systems	Senior	41....45	66k....70k	3
Marketing	Senior	36....40	46k....50k	10
Marketing	Junior	31....35	41k....45k	4
Sales	Senior	31....35	46k....50k	30
Sales	Junior	26....30	26k....30k	40
Sales	Junior	31....35	31k....35k	40

- 9 Consider the following data points: (7)

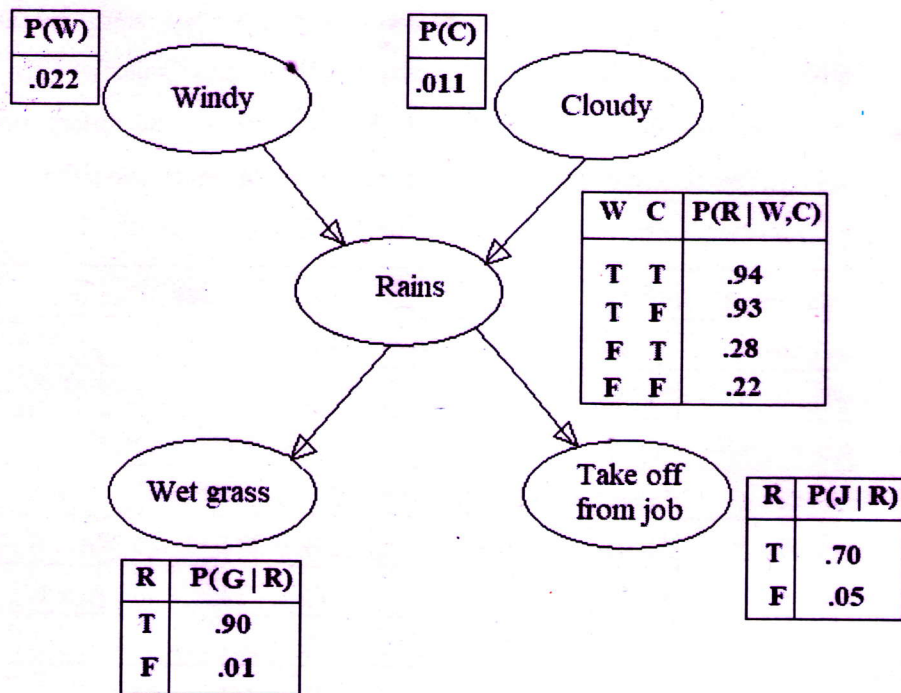
A1(3, 9), A2(3, 4), A3(9, 4), A4(6, 8), A5(8, 5), A6(7, 4), A7(3, 2).

Identify the cluster centers by applying the k-means algorithm, with $k = 2$.

Initially assign A1 and A4 as the clusters center respectively.

The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as $D(a, b) = |x_2 - x_1| + |y_2 - y_1|$

- 10 Shown below is the Bayesian network corresponding to the Wet grass problem, $P(J | R) P(G | R) P(R | W, C) P(W) P(C)$. The probability tables show the probability that variable is True ($P(M)$ means $P(M = T)$). Calculate the joint probability of having wet grass, given it has not rained and it is windy and not cloudy. (7)



- 11 Given the following data, construct the Receiver Operator Characteristic (ROC) curve of the data. Compute the AUC. (7)

Threshold	TP	TN	FP	FN
1	0	5	0	5
2	1	5	0	4
3	1	4	1	4
4	3	4	1	2
5	3	3	2	2
6	4	3	2	1
7	4	2	3	1
8	4	1	4	1
9	5	0	5	0

- 12 Identify and explain any three model combination approach to improve the accuracy of a classifier. (7)
